1-1-2006

Human genomic deletions mediated by recombination between Alu elements

Shurjo K. Sen
Louisiana State University

Kyudong Han
Louisiana State University

Jianxin Wang
Roswell Park Cancer Institute

Jungnam Lee
Louisiana State University

Hui Wang
Louisiana State University

See next page for additional authors

Recommended Citation

Sen, S., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P., Dyer, M., Cordaux, R., Liang, P., & Batzer, M. (2006). Human genomic deletions mediated by recombination between Alu elements. *American Journal of Human Genetics, 79* (1), 41-53. https://doi.org/10.1086/504600

Human Genomic Deletions Mediated by Recombination between *Alu* Elements

Shurjo K. Sen,* Kyudong Han,* Jianxin Wang, Jungnam Lee, Hui Wang, Pauline A. Callinan, Matthew Dyer, Richard Cordaux, Ping Liang, and Mark A. Batzer

Recombination between *Alu* elements results in genomic deletions associated with many human genetic disorders. Here, we compare the reference human and chimpanzee genomes to determine the magnitude of this recombination process in the human lineage since the human-chimpanzee divergence ~6 million years ago. Combining computational data mining and wet-bench experimental verification, we identified 492 human-specific deletions (for a total of ~400 kb) attributable to this process, a significant component of the insertion/deletion spectrum of the human genome. The majority of the deletions (295 of 492) coincide with known or predicted genes (including 3 that deleted functional exons, as compared with orthologous chimpanzee genes), which implicates this process in creating a substantial portion of the genomic differences between humans and chimpanzees. Overall, we found that *Alu* recombination-mediated genomic deletion has had a much higher impact than was inferred from previously identified isolated events and that it continues to contribute to the dynamic nature of the human genome.

With a copy number of >1 million, *Alu* elements are one of the most successful non-LTR (long terminal repeat) retrotransposon families in the human genome.[1] In addition to classic retrotransposition-associated insertion mutations, *Alu* elements can create genomic instability by the deletion of host DNA sequences during their integration into the genome and by creating genomic deletions associated with intrachromosomal and interchromosomal recombination events.[2,3] Multiple features predispose *Alu* elements to successful recombination, including their proximity in the genome (one insertion every 3 kb, on average), the high GC content of their sequence (~62.7%), and the remarkable sequence similarity (70%–100%) among *Alu* subfamilies of widely different ages. Overall, the recombinogenic nature of these elements is reflected in the various forms of cancer and genetic disorders associated with *Alu*-mediated recombination events.[3–12]

However, clinical studies of isolated disease-causing deletions, although useful from a medical viewpoint and in demonstrating the existence of *Alu* recombination-mediated deletions (ARMDs), do not adequately depict the overall contribution of this process to the architecture of the genome and the associated impact on gene function. The availability of a genome sequence for the common chimpanzee (*Pan troglodytes*), the closest evolutionary relative of the human lineage,[13] has allowed us to perform a comparative genomic assessment of the extent of ARMD in the human genome over the past ~6 million years, since the divergence of the human and chimpanzee lineages.[14,15] In this study, we identified ~400 kb of human-specific ARMD, the distribution of which is biased toward gene-dense regions of the genome, which raises the possibility that ARMD may have played a role in the divergence of humans and chimpanzees. About 60% of the ARMDs are located in genes, and, in at least three instances, exons have been deleted in human genes relative to their chimpanzee orthologs. The nature of the altered genes suggests that ARMD might have played a role in shaping the unique traits of the human and chimpanzee lineages. Mechanistically, we characterized the physical aspects of the deletion process and proposed different models for ARMD.

Material and Methods

Computational Data Mining for Identification of Candidate ARMD Loci

We extracted 400 bp of 5′ and 3′ genomic sequence flanking all human *Alu* elements (fig. 1). Next, we joined the two 400-bp stretches to form a single sequence (the "query"). For each query, the best match in the reference chimpanzee genome (PanTro1 [November 2003 freeze]) was identified. Then, the sequence stretch in the chimpanzee genome between the two regions that aligned with the two 400-bp halves of the query (the "hit") was extracted and aligned with the human *Alu* sequence initially used to design the query (the "query *Alu*"), by use of a local installation of the National Center for Biotechnology Information Blast 2 Sequences Bl2seq utility. Following are the possible alignment results for each sequence pair (see corresponding diagrams in fig. 1).

A. There is no match. In this case, an *Alu* insertion-mediated deletion has occurred in the human genome at that locus.
B. There is only one alignment block, and:

From the Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, Baton Rouge (S.K.S.; K.H.; J.L.; H.W.; P.A.C.; M.D.; R.C.; M.A.B.); and Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo (J.W.; P.L.)

Received February 6, 2006; accepted for publication March 22, 2006; electronically published May 3, 2006.

Address for correspondence and reprints: Dr. Mark. A. Batzer, Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803. E-mail: mbatzer@lsu.edu

* These two authors contributed equally to this work.

Am. J. Hum. Genet. 2006;79:41–53. 0002-9297/2006/7901-0006$15.00

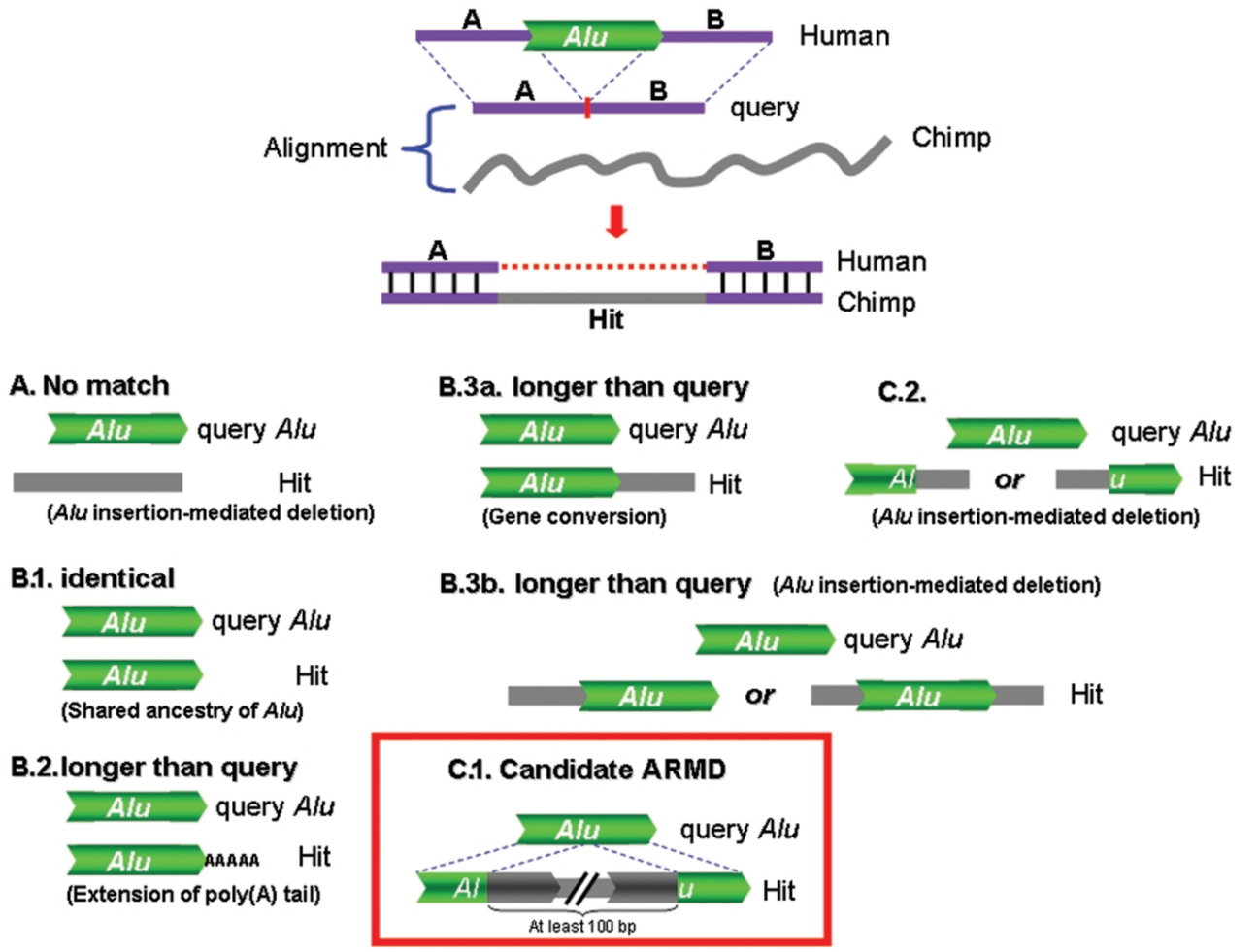

Figure 1. Computational data mining for human lineage-specific ARMD loci. *A,* No match between query *Alu* and hit (possible *Alu* insertion–mediated deletion). *B.1,* Query *Alu* and hit are identical (shared ancestry of an *Alu* insertion). *B.2,* Hit is longer than query *Alu and* the extra sequence is a poly(A) tract downstream of the query *Alu* (extension of the *Alu* poly(A) tail). *B.3,* Hit consists of query *Alu* plus extra non-poly(A) sequence, *and* the following. *B.3a,* extra, non-poly(A) sequence is downstream of the query *Alu* poly(A) tail (may be gene conversion event in the chimpanzee genome). *B.3b,* Extra, non-poly(A) sequence is upstream of the query *Alu* element or there is extra sequence at both ends (possible *Alu* insertion–mediated deletion event). *C.1,* Beginning and end of the hit match query *Alu and* the hit is at least 100 bp longer than query *Alu* (candidate human lineage-specific ARMD event). *C.2,* At least one end of the hit has no match to query *Alu* (possible *Alu* insertion–mediated deletion).

B.1. The hit is identical to the query *Alu.* This is shared ancestry of an *Alu* insertion.

B.2. The hit is longer than the query *Alu,* and the extra sequence is entirely composed of a poly(A) tract downstream of the *Alu* sequence. This is a case of extension of the *Alu* poly(A) tail.

B.3. The hit consists of the query *Alu* plus some extra non-poly(A) sequence, and

B.3a. The extra, non-poly(A) sequence is downstream of the poly(A) tail. This could be a gene conversion event in the chimpanzee genome.

B.3b. The extra, non-poly(A) sequence is upstream of the query *Alu* element or there is extra sequence at both ends. This is a possible *Alu* insertion–mediated deletion event in the human genome.

C. There is more than one alignment block, and

C.1. The beginning and end of the hit match the query *Alu and* the hit is at least 100 bp longer than the query *Alu* sequence (since this size would approximate the expected lower ARMD size limit). This is a candidate ARMD event in the human genome.

C.2. At least one end of the hit has no match to the query *Alu.* This is another possible case for an *Alu* insertion–mediated deletion in the human genome.

We retained all loci matching case C.1 as pairs of FASTA files (i.e., the orthologous human and chimpanzee sequences). Each human sequence contained the query *Alu* and its 400-bp flanking sequences on each side, and each chimpanzee sequence contained the entire hit that aligned with the query flanking sequences. All candidate ARMD loci were then manually inspected and, if necessary, verified by wet-bench (PCR) analysis. Orthologous human and chimpanzee sequences for each locus are available from the "Publications" section of the Batzer Laboratory Web site.

Inspection of Target-Site Duplications

A typical *Alu* insertion is flanked on both sides by identical (or nearly perfect) short, direct repeats (7–20 bp) termed "target-site duplications" (TSDs).[16] The single *Alu* element remaining at a human candidate ARMD locus is characterized by the apparent absence of TSDs, since it is composed of fragments from a pair of *Alu* elements with mutually different TSDs, situated at the orthologous ancestral locus (which persists in the chimpanzee genome). This hallmark of the ARMD process offers a direct means of confirming the "chimeric" origin of the human *Alu* element at a deletion locus. Using this property as our basis for verification, we manually inspected all candidate loci returned by the computational analysis. In an unambiguous ARMD event, the TSDs of the two *Alu* elements immediately upstream and downstream of the deleted portion in the chimpanzee genome were perfect matches with the 5′ and 3′ TSDs, respectively, of the orthologous single human *Alu* element. In the next possible scenario, the sequence on any one side of the human *Alu* element (upstream or downstream) matched the TSDs of the chimpanzee element on the corresponding side, but the other chimpanzee *Alu* element itself lacked TSDs. However, the sequence immediately flanking this element on the side opposite to the deletion was identical in both human and chimpanzee. In both these cases, we accepted the computational detection as a valid ARMD locus. At loci that showed slight deviations in the sequence architecture from the unambiguous ARMD structures described above (which raise the possibility that one of the two chimpanzee *Alu* elements might be a chimpanzee-specific *Alu* insertion, as opposed to a human-specific ARMD event), we designed oligonucleotide primers in the nonrepetitive sequences flanking the *Alu* elements in the chimpanzee genome, and we experimentally confirmed by PCR (and, where required, by DNA sequencing) that the deletion did exist and was specific to the human genome.

As an additional step to verify the potential ARMD loci that we accepted/rejected solely on the basis of computational identification, we randomly chose two sets of 25 such insertions and deletions and verified them by PCR. Accuracy rates for putative deletion and insertion loci were 100% and 96%, respectively (4% of putative insertions comprising the error were all deletions), which confirmed the validity of our approach.

PCR Amplification and DNA Sequence Analysis of ARMD Loci

We designed oligonucleotide primers using Primer3 software. Detailed information for each locus—including primer sequences, annealing temperature, and PCR product sizes—is available from the "Publications" section of the Batzer Laboratory Web site.

PCR amplification of each locus was performed in 25-μl reactions with 10–50 ng genomic DNA, 200 nM of each oligonucleotide primer, 200 μM dNTPs in 50 mM KCl, 1.5 mM MgCl2, 10 mM Tris-HCl (pH 8.4), and 2.5 units *Taq* DNA polymerase. The conditions for the PCR were an initial denaturation step of 94°C for 4 min; followed by 32 cycles of 1 min of denaturation at 94°C, 1 min of annealing at optimal annealing temperature, and 1 min of extension at 72°C; followed by a final extension step at 72°C for 10 min. PCR amplicons were separated on 2% agarose gels, were stained with ethidium bromide, and were visualized using UV fluorescence.

Individual PCR products were purified from the gels with Wizard gel purification kits (Promega) and were cloned into vectors by use of TOPO-TA Cloning kits (Invitrogen). For each sample, three colonies were randomly selected and were sequenced on an Applied Biosystems ABI3130*XL* automated DNA sequencer. Each clone was sequenced in both directions with use of M13 forward and reverse primers. The sequence tracks were analyzed using the Seqman program in the DNASTAR suite and were aligned using BioEdit sequence alignment software. Gorilla and orangutan sequences generated during the course of this study have been submitted to GenBank under accession numbers DQ363502–363524.

Loci verified by PCR were screened on a panel of five primate species, including *Homo sapiens* (HeLa; cell line ATCC CCL-2), *P. troglodytes* (common chimpanzee; cell line AG06939B), *Pan paniscus* (bonobo or pygmy chimpanzee; cell line AG05253B), *Gorilla gorilla* (western lowland gorilla; cell line AG05251), and *Pongo pygmaeus* (orangutan; cell line ATCC CR6301). To evaluate polymorphism rates, we amplified 50 randomly picked ARMD loci on a panel of genomic DNA, from 80 human individuals (20 from each of four populations: African American, South American, European, and Asian), that was available from previous studies in our lab.

Monte Carlo Simulations of GC and Alu Content

To test whether the GC and *Alu* contents of the sequences deleted through ARMD differed statistically from the rest of the genome, we performed Monte Carlo simulations comparing the observed deletions to two other sets of sequences. Both these sets comprised randomly extracted sequences equal in number to the observed deletions (492) and mimicked the observed size distribution of ARMD events. The first set was extracted from the regions immediately adjacent to randomly picked *Alu* elements annotated in the reference human genome sequence (called "RSNA" hereafter). The second set comprised sequences randomly extracted from the entire genome sequence, with no additional parameters incorporated (called "RSG" hereafter). We used 5,000 randomized replicates of both sets. For both observed and simulated sets of sequences, we calculated GC content using in-house Perl scripts, whereas the *Alu* content was analyzed using a locally installed copy of the RepeatMasker Web server. Additionally, to make our estimate of observed percentage *Alu* content conservative, we trimmed the deleted sequence at each locus to remove remaining fragments of the two *Alu* elements that caused the ARMD event.

Statistical significances of the differences in GC and *Alu* content were based on *Z* scores obtained by comparing observed values (from the actual set of deleted sequences) with the mean value obtained from the 5,000 randomly extracted sequence sets.[17] All computer programs used are available from the authors on request.

Results

A Whole-Genome Analysis of Human-Specific ARMD Events

To identify putative ARMD loci, we first computationally compared the human and chimpanzee genomes. Subsequently, we manually inspected and, if needed, experimentally verified individual loci. Of the 1,332 computationally predicted deletions that we initially recovered, 461 were discarded after manual inspection (table 1). The causes for rejection of computationally predicted ARMD loci were: (*a*) insertion of an *Alu* or other retroelement at the orthologous chimpanzee locus, which leads to the presence of sequence that the computer erroneously as-

sumed to be deleted in the human genome (38 cases), (*b*) authentic deletion products in the human genome that were not products of *Alu-Alu* recombination (211 cases), and (*c*) computational errors in alignment of the human and chimpanzee genomes (212 cases). On the basis of sequence architecture, the remaining 871 loci represented putative ARMD events in the human lineage. All of these loci were further manually inspected and were analyzed, for comparison of the ancestral predeletion and human postdeletion states, by use of a TSD-based strategy as described above (see the "Material and Methods" section). In addition, we experimentally verified the authenticity of 352 candidate ARMD loci by PCR (table 1 and fig. 2). To be conservative, we discarded all loci in which an alternative mechanism (e.g., random genomic deletion), distinct from ARMD, could have produced the deletion. Specifically, ARMD events can be distinguished from random genomic deletions occurring at *Alu* insertion sites because an ARMD event reconstitutes an uninterrupted chimeric *Alu* element (i.e., with no internal deletion), whereas the probability of this happening through chance alone (as would be the case with a random deletion) is remote. Indeed, the probability of two ~280-bp *Alu* elements breaking by chance at a homologous site is only 1 in ~80,000 (1 in 280 × 1 in 280). Hence, although we cannot formally exclude the possibility that a few random deletions may precisely mimic the ARMD process, we believe the overall impact of these nonauthentic events on our estimates would be minimal.

Table 1. Summary of Human-Specific ARMD Events

Classification	No. of Loci
Computationally predicted deletion loci	1,332
Discarded after manual inspection	461
Candidate ARMD events:	871
False-positive events (*Alu* insertion in chimpanzee):	379
Confirmed by PCR analysis	189
Analysis based on TSD structure	190
ARMDs:	492
Confirmed by PCR analysis	163
Analysis based on TSD structure	329

The manual verification of the 871 loci resulted in a final data set of 492 ARMD events spanning the entire human genome (table 1). Nine ARMD loci on the Y chromosome were all located in the pseudoautosomal part of this chromosome and hence were identical copies of deletion loci on the X chromosome. As a result, each event was counted only once during the analysis. In general, the loci analyzed in this study suggest that the combination of computational data mining and experimental validation is the "gold standard" when conducting comparative genomic searches for lineage-specific deletions. As we observed during the course of this study, lineage-specific in-

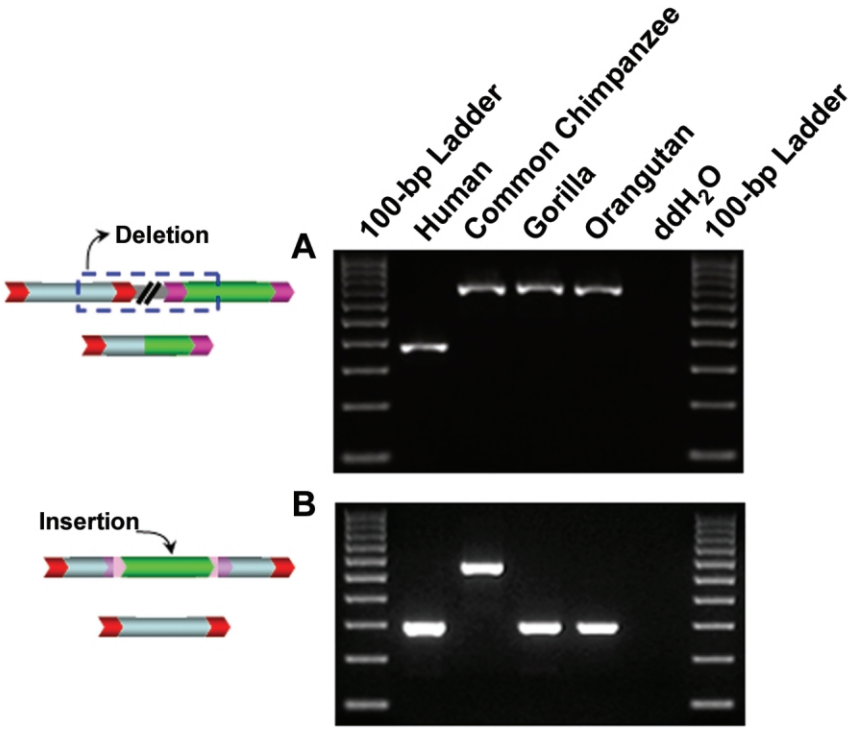

Figure 2. ARMD in the human genome: individual ARMD candidate loci amplified by PCR. *A,* Agarose-gel chromatograph of PCR products derived from an authentic human-specific ARMD event. *B,* Agarose-gel chromatograph of PCR products derived from an ARMD false-positive event (*Alu* insertion in chimpanzee). The DNA templates used in each reaction are shown above the chromatographs.

The figure is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Figure 3. Density of ARMD events and all *Alu* insertions on individual human chromosomes.

sertions in one genome stand a risk of being characterized as deletions in the other when only two genomes are compared in a computational analysis. In our analysis, we minimized the chances of including such events by using three other hominoid genomes as controls during experimental verification of the events.

Extent of Genomic Deletion and Size Distribution of ARMD Events

The number of ARMD events is positively correlated with the number of *Alu* elements present on each chromosome ($r = 0.69$; $P < .0005$). This is expected, since physical proximity between repetitive elements strongly predisposes them to recombination.[18] Simultaneous mapping of ARMD loci and all *Alu* insertions on each chromosome highlights the tendency for deletions to cluster with regions of high local *Alu* density (fig. 3). Additionally, sequence analysis of the *Alu* elements involved in ARMD events indicates that the number of elements from each *Alu* subfamily (fig. 4) is proportional to their genomewide copy number,[6] with no bias observed for elements from older subfamilies (such as *Alu*J) that would have had more time for recombination because of their age. This implies that *Alu* elements throughout the genome have similar chances of recombining with each other, as opposed to a mechanism of preferential recombination between members of an individual subfamily, and that proximity between the elements is the major factor involved in the process. Additional evidence supporting this position comes from the fact that ~40% (197 of 492) of ARMD events result from inter–*Alu* subfamily recombinations. However, within this context, the amount of sequence identity between the two elements at a locus also appears to be proportional to their chances of successful recombination, since young *Alu*Y elements are overrepresented at ARMD loci compared with their total number in the genome, whereas the opposite is true for older, highly diverged *Alu*J elements.

The total amount of genomic sequence deleted by this process in the human lineage (i.e., after the human-chimpanzee divergence ~6 million years ago) is estimated to be 396,420 bp. This is probably a conservative estimate, since our comparative analysis of the human and chimpanzee genomes detects ARMD events only between *Alu* elements that were inserted before the human-chimpanzee divergence. Therefore, it would miss ARMD loci involving newly inserted human-specific *Alu* elements.[19,20] However, the contribution of human-specific *Alu* elements to ARMD is probably relatively limited, given that there are only ~7,000 such insertions,[13] as compared with >1 million *Alu* elements shared between the human and chimpanzee genomes.

The ARMDs range in size between 101 bp and 7,255 bp, with an average size of ~806 bp. A histogram of the size frequency distribution of ARMDs reveals a skew toward shorter ARMD sizes, with ~75% (368 of 492) of the deletions shorter than 1 kb (fig. 5). Thus, the median ARMD length of 468 bp better represents the most common size category. However, in terms of total genomic sequence deleted, the ~25% ARMD events >1 kb were responsible for ~62% (245,263 of 396,420 bp) of the total sequence deleted. Our computational analyses did not return any ARMD loci with deletions <100 bp. Strictly speaking, *Alu-Alu* recombination elements should not cause deletions of <300 bp (i.e., the length of a complete *Alu* element), because, even if the recombining elements were immediately adjacent to each other, this would be the smallest possible amount of sequence deleted. However, the individual left and right monomers of the dimeric *Alu* element can freely exist in the genome, and these types of elements are accounted for in our study. This resulted in

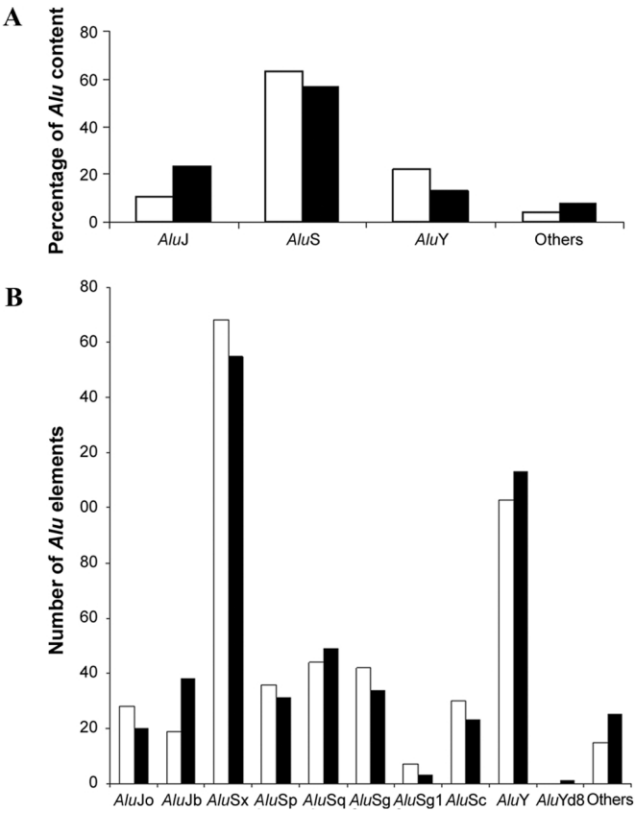

Figure 4. *Alu* subfamily composition in ARMD events. *A,* Proportion of *Alu* elements involved in ARMD events (*unblackened bars*) versus total number of *Alu* elements (*blackened bars*) for each subfamily. *B,* Subfamily ratios of upstream and downstream *Alu* elements involved in ARMD events (unblackened and blackened bars, respectively).

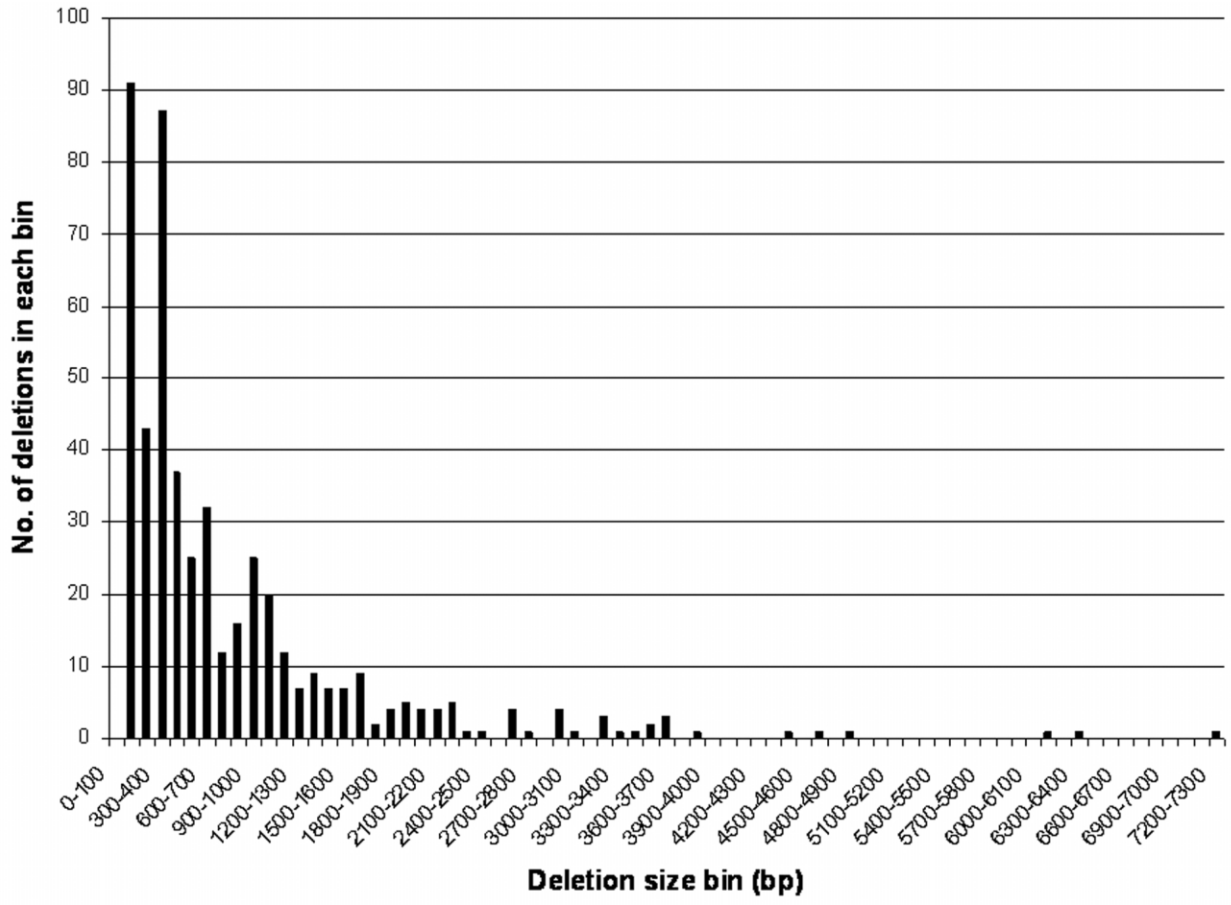

Figure 5. Size distribution of human-specific ARMD events, displayed in 100-bp bin sizes

the ability of our study to detect deletions smaller than the expected minimum of ~300 bp.

Structural Characteristics of ARMD Events

Pairs of *Alu* elements that recombined to cause human genomic deletions were in parallel orientation in almost all cases (490 of 492). Most probably, this is a direct consequence of the increased length of hybridization available from this arrangement, since the parallel orientation would allow for homology over longer stretches between pairs of *Alu* elements located on the homologous chromosomes during recombination. Analysis of the *Alu* trios at each locus (i.e., two pre-ARMD *Alu* elements in chimpanzee and one postdeletion element in human) suggests four possible recombination mechanisms. Of these, unequal recombination between adjacent *Alu* elements on homologous chromosomes (fig. 6*A*, left panel) accounts for ~74% (366 of 492) of the deletions, whereas the other three putative mechanisms were less frequent (fig. 6*B*–6*D*). Our study captures both intrachromosomal (fig. 6*A*, right panel) and interchromosomal (fig. 6*A*, left panel) recombination-mediated deletions.

For each deletion, we located the points on the *Alu* consensus sequence where the two intact chimpanzee *Alu* elements involved in the recombination were broken and subsequently attached to each other to form the resulting single human *Alu* element. Plotting the frequency distribution of recombination breakpoints at different positions on the *Alu* consensus sequence revealed a recombination "hotspot" encompassing positions 21–48 (fig. 7), which is consistent with an earlier study based on a smaller data set.[21] To uncover the reasons underlying the observed "adhesive" nature of this part of the *Alu* element, we aligned the consensus sequences of 10 *Alu* subfamilies (*Alu*Jo, *Alu*Jb, *Alu*Sx, *Alu*Sp, *Alu*Sq, *Alu*Sg, *Alu*Sg1, *Alu*Sc, *Alu*Y, and *Alu*Yd8) and analyzed the levels of conservation and GC content of regions that tended to recombine at frequencies exceeding the mean (0.08) across all positions in our ARMD events. This analysis indicated that both parameters were substantially higher in these regions than in the rest of the *Alu* sequence, with the major inferred recombination hotspot referred to above showing >60% GC (as compared with the ~62.7% average GC content for the 10 *Alu* consensus sequences) and complete conservation across all subfamilies. Although these factors may be responsible for higher recombination frequencies in this region, other

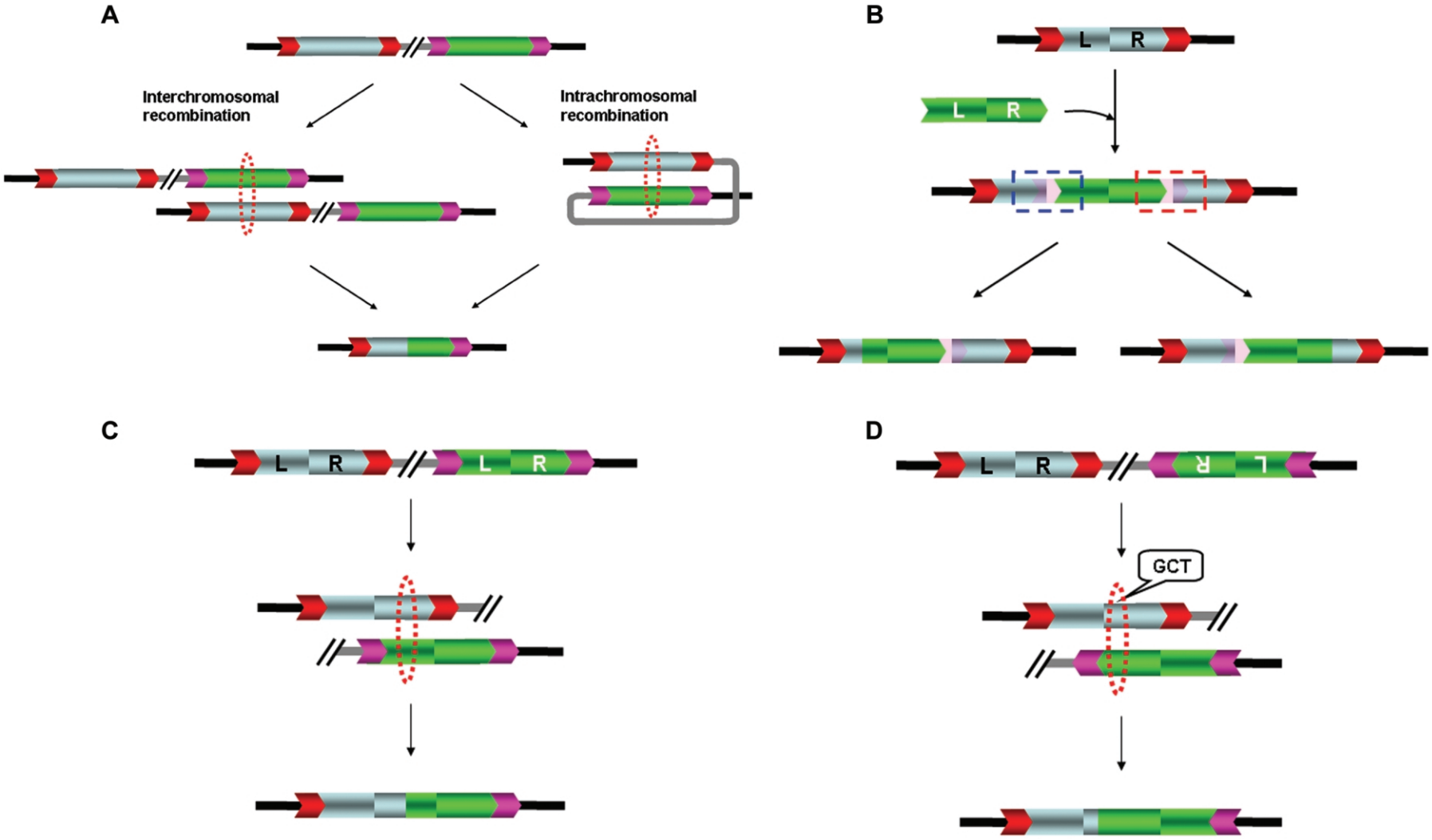

Figure 6. Four different types of recombination between *Alu* elements. Black and gray lines represent flanking and intervening regions, respectively. Dotted red circles denote recombining regions, and red and pink arrows represent TSDs of the two elements, respectively. *A,* Interchromosomal (*left*) and intrachromosomal (*right*) recombination between two *Alu* elements (*light blue* and *green*). *B,* Recombination between two *Alu* elements, one of which previously inserted into the other (L and R indicate left and right *Alu* monomers). *C,* Recombination between left and right *Alu* monomers on two different elements. *D,* Recombination between oppositely oriented *Alu* elements (only two cases observed).

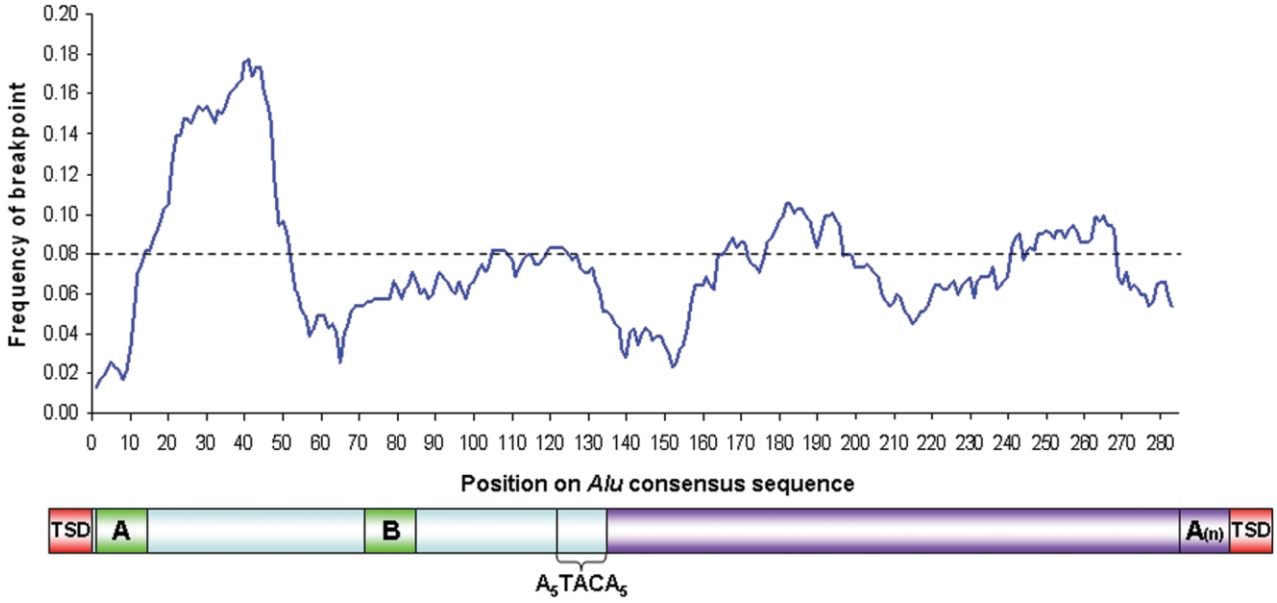

Figure 7. Recombination window between *Alu* elements and percentage frequencies of breakage (during recombination) at different positions along an *Alu* consensus sequence. The structure of a typical *Alu* element is shown in the lower panel. The length of the *Alu* consensus sequence is ~282 bp, excluding the 3′ poly(A) tail. The element consists of left (*light blue*) and right (*purple*) monomers. The left monomer contains an RNA polymerase III promoter (*green boxes A* and *B*). TSDs (*red boxes*), usually 7–20 bp long, are created at each end during the *Alu* insertion process.

reasons are also plausible, such as the location of this stretch near the L1 endonuclease cleavage site at the 5′ end of the *Alu* element, which makes it closer to putative breakage sites during the recombination process.

Genomic Environment of ARMD Events

Alu elements in the human genome show a preference for high–GC content areas, except for the most recently integrated subfamilies.[1,22] However, since only a fraction (984 of ~1.2 million) of the total number of *Alu* insertions is associated with the ARMD process, it may well be that, in this respect, the deletions themselves behave differently from the *Alu* family as a whole. To characterize the sequence context in which ARMD events occur, we calculated the percentage of GC content in 20-kb windows of flanking sequence centered on the ARMD loci. Compared with previous analyses of *Alu* and L1 insertion-mediated (as opposed to postinsertional recombination–mediated) genomic deletions,[2,23] which are preferentially localized in low–GC content neighborhoods (~38% GC), ARMD events tend to occur in high–GC content regions (~45% GC content, on average). This is also substantially higher than the ~41% global average GC content of the human genome.[1] Since high–GC content areas of the genome also show higher gene density,[1,24] we analyzed 4 Mb (2 Mb in each direction) of sequence flanking ARMD events, for the presence of known and predicted human RefSeq genes. We found the gene density around ARMD events to be, on average, one gene per 66 kb, which, as expected, is higher than the global average gene density (approximately one gene per 150 kb)[24] and the average gene density in the vicinity of L1 insertion-mediated deletions (approximately one gene per 200 kb).[23] Thus, ARMD events seem to be concentrated in gene-rich regions of the human genome. The tendency for clustering of ARMD events and genes becomes even more apparent when their densities are plotted side by side on each chromosome (fig. 8). Interestingly, the neighboring GC content showed a significant negative correlation with the deletion size ($r = -0.17$; $P < .0001$).

About 45% (219 of 492) of ARMD events were located within known or predicted human RefSeq genes, and an additional ~15% (76 of 492) were in intergenic regions of the human genome but were located within predicted chimpanzee genes. Since ⩽25% of the human genome represents currently known genes (including both exon and intron sequences),[24–26] the relative density of ARMD events within genic regions is remarkably high. This would indicate that, a priori, the probability of this process interfering with gene function is higher than the two retrotransposon insertion-mediated deletion mechanisms mentioned above. To test this hypothesis, we extracted the ancestral prerecombination sequence at each ARMD locus (i.e., the sequence present in the chimpanzee genome but deleted in the human genome) and analyzed its location in the chimpanzee genome to see whether it mapped to a protein-coding region. In three instances, the ARMD event deleted an entire exon from a gene that is

The figure is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Figure 8. Density of ARMD events and RefSeq genes on individual human chromosomes.

functional in the chimpanzee genome. To confirm that these three ARMD loci did not represent assembly errors, we resequenced them in the human and chimpanzee genomes. One of the three genes, *LOC471177,* is a model chimpanzee gene similar to the human *CHRNA9* gene (MIM 605116), a member of the ligand-gated ionic channel family that is associated with cochlea hair cell development.[27] Of the other two, *LOC452742* is similar to the human model gene *LOC440141* (which encodes the mitochondrial ribosomal protein S31), and *LOC471116* encodes a hypothetical protein with a conserved high–molecular weight glutenin subunit.

Characteristics of the Genomic Sequences Lost during ARMD

Previous analyses have suggested that recombination may be responsible for the bias toward high–GC content areas observed for *Alu* elements in the human genome.[1,28–30] If so, one would expect that ARMD events preferentially remove low–GC content sequence, consequently causing a shift in the opposite direction. However, simulation results revealed that the GC content of both RSNA and RSG (41.9% and 41.4%, respectively) were significantly lower than the ~45.4% GC content of the observed deleted sequences ($P < .00001$ in both cases). Moreover, the RSNA and RSG *Alu* contents (20.6% and 11.4%, respectively) also had significantly lower values when compared with the *Alu* content of the observed deleted sequences (27.0%; $P < .0001$, compared with both RSNA and RSG). In addition to *Alu* elements, repetitive DNA from elements of other families, for a total of 86,442 bp, was removed by ARMD (table 2).

Discussion

Role of the ARMD Process in Human Genome Evolution

Retrotransposons such as *Alu* elements are associated with size expansion in primate genomes.[31,32] This is a consequence of their increasing copy number and also an indirect result of their implication in homology-mediated segmental duplications.[33] For example, the high retrotransposition activity of the *Alu* family in the human lineage has been responsible for the addition of ~2.1 Mb to the human genome within the past ~6 million years.[13,34] In this context, our study provides the first comprehensive assessment of a postretrotransposition process that has had an appreciable impact on the dynamics of human genome–size evolution. Previous in vivo evolutionary analyses have characterized human and chimpanzee genomic deletions generated on *Alu* and L1 insertions[2,23] However, the combined extent of human-specific deletion attributable to these mechanisms is an order of magnitude lower than that resulting from ARMD (~30 kb for *Alu* and L1 insertion–mediated deletions combined, vs. ~400 kb for ARMD alone). The relative amounts of sequence inserted (by *Alu* retrotransposition) and deleted (by ARMD) imply an *Alu*-mediated sequence turnover rate of ~20% (i.e., ~400-kb deleted sequence vs. ~2.1-Mb inserted sequence) in the human genome within the past ~6 million years. This indicates that ARMD is capable of mitigating, at least partially, the increase in genome size caused by new retrotransposon insertions.

The scope of retrotransposon-mediated reduction of genome size further broadens when we consider that L1 elements (another mobile DNA family) are capable of creating deletions by a recombination process analogous to ARMD.[33,35] The higher average distance between L1 insertions in the human genome (one element per 6.3 kb)[1] as well as the lower GC content of L1 elements (~43%, excluding the poly(A) tail)[36] may be contributing factors to the paucity of L1-mediated recombination events as compared with ARMD events. Even so, the greater length of L1 elements (~6 kb vs. ~300 bp for *Alu* elements)[36] and their high copy number (~520,000 elements)[1] still indicate that this family may represent another source of retrotransposon recombination-mediated deletions in the human genome. However, a broader comparative genomic study of such retrotransposon recombination-mediated deletion mechanisms in both the human and chimpanzee lineages is needed before the comprehensive role of transposable elements in primate genome–size evolution can be determined. In this respect, at least in the case of plants, studies have already shown that the genome of *Arabidopsis thaliana* uses recombination-mediated deletion to counterbalance genome expansion, which may be one of the reasons for its remarkably compact size.[37]

Recent analyses of human-genome variation have emphasized the importance of deletions in creating genetic diversity among humans.[38–41] Our results offer insight into one of the mechanisms that may contribute to the crea-

Table 2. Genomic DNA Sequences Deleted by ARMD

Classification	Amount
Alu[a]	192,102
MIR	4,780
7SL RNA	306
L1	41,491
L2	7,312
L3	163
LTR	23,336
MER1	3,575
MER2	2,555
Other DNA repeat elements	669
Simple repeat	2,255
Nonrepetitive DNA	117,876
Total	396,420

[a] Includes truncated *Alu* elements.

tion of such deletions. Interestingly, the majority of the deletion variants identified in the recent studies cited above[39–41] are polymorphic between human individuals or populations. Although their contribution to between-individual genetic diversity is undisputed, the persistence of these deletions over evolutionary time cannot be taken for granted. By contrast, the deletions reported in our study have a low polymorphism rate (15%) among the 80 diverse human genomes we genotyped. This may represent the difference in the comparative time scales of these between-human genomic deletion variants[39,40] and our human-chimpanzee comparison. In an earlier analysis,[23] we showed that only a fraction of the deletions caused by in vitro L1 retrotransposition[42–44] persist in the human genome over evolutionary time. Additionally, comparative genomic studies across a range of organisms indicate that genomic deletions that ultimately reach fixation tend to be smaller than those detected before any selective force operates (i.e., in cell culture analyses).[45] Analogous to this situation, ARMD events (which had a median length of 468 bp) were, in general, smaller than the deletion variants characterized by the recent studies of human-genome variation, which had a range of 1–745 kb.[39–41] Since our study focuses on a longer evolutionary time scale and would preferentially capture those ARMD events that have not been selected against, it is possible that the deletions we detected represent the smaller evolutionary remainder of a group of older and perhaps larger deletions.

ARMD as an Agent in Human-Chimpanzee Divergence

The human and chimpanzee genomes are characterized by only ~1.4% divergence at the nucleotide-sequence level.[13,46–48] With the completion of the draft chimpanzee genome, the focus has shifted to identifying differences rather than locating similarities. Regarding actual genetic change, although a comprehensive assessment of protein-coding portions of the chimpanzee genome is not yet available, functional classes of genes that are under accelerated evolution in one lineage or the other have been characterized by recent studies.[49,50]

In the context of possible events that have altered gene structure or expression between the human and chimpanzee lineages, our study illustrates almost 300 lineage-specific deletions within protein-coding human or chimpanzee RefSeq genes; it is conceivable that at least some of these ARMD events contributed to phenotypic divergence. Gene shuffling by recombination between *Alu* elements has already been reported in the human genome.[51] Furthermore, in at least two documented instances, *Alu* elements have caused hominoid lineage-specific exon deletions in functional genes: through an insertion-mediated deletion in the human *CMAH* gene[52] and through ARMD in the human *ELN* gene.[53] In the present study, we show three additional instances in which ARMD has caused the loss of an exon in a human gene, as compared with its chimpanzee ortholog. Of particular interest is the deletion of the fourth exon in the predicted chimpanzee gene *LOC471177*, which is orthologous to the human *CHRNA9* gene. In the human lineage, CHRNA9 is an ionotropic receptor with a probable role in the modulation of auditory stimuli.[27,54] Modifications in the function of this gene may lead to a reduction in basilar membrane movement and thus affect the dynamic range of hearing. Although the characterization of the actual gene expression pathways that underlie the differences of humans and chimpanzees has just begun, preliminary data suggest that differences in auditory genes may comprise a subset of the total change.[49] This is reflected in the fact that the tonal range of normal human speech is probably outside the optimal reception of the chimpanzee auditory system.[55] Thus, it is conceivable that *CHRNA9* is a member of the group of genes (such as *FOXP2* and *TECTA*) that may be responsible for the unique auditory and olfactory traits that distinguish humans and chimpanzees.[49,56] Even excluding the three ARMD events listed above that deleted exons, 292 other events located within genes have deleted 229,205 bp of intronic sequence. Although further analysis will be required for conclusive assignment of specific roles, if any, to the deleted intronic sequences, it is possible that some of them may be associated with alteration of splicing patterns.

Does ARMD Play a Role in Modifying Alu *Distribution?*

Recently integrated or young *Alu* elements are inserted relatively randomly in the genome; by contrast, older *Alu* elements are preferentially found in GC-rich areas of the genome.[1,22] Both selective and neutral explanations have been offered for this uneven genomic distribution of *Alu* elements. However, a selective process[1] is inconsistent with polymorphism patterns of recently integrated *Alu* elements.[22] An alternative explanation for the enrichment of *Alu* elements in GC-rich regions over time involves their preferential loss from GC-poor regions,[1,28–30] a process that might be influenced by ARMD.

However, the high GC content of deleted sequences, along with the preferential occurrence of ARMD events in GC-rich regions, argues against this possibility. To result in the *Alu* distribution shift, the deletions would need to be much larger in GC-poor than in GC-rich regions.[22] Consistent with this hypothesis, our results indicate that ARMD size is negatively correlated with GC content. However, although ARMD events are significantly larger in GC-poor (i.e., <41% genome average) than in GC-rich (i.e., >41% genome average) regions (~1,100 vs. ~700 bp; t test $P = .0007$), three times as many ARMD events occurred in GC-rich as in GC-poor regions (369 vs. 123). Consequently, the net amount of sequence deleted from GC-poor regions is half that of GC-rich regions (~135 kb vs. ~261 kb). Given that GC-poor regions encompass ~58% of the genome,[1] it is unlikely that ARMD has played a substantial role in mediating the shift in the *Alu* distribution toward heavy isochores.[13] Nevertheless, other types of deletions could contribute more significantly to the yet-unexplained *Alu* genomic distribution shift.

Interestingly, the results from the simulations we performed suggest that sequences deleted through ARMD contain a statistically significant excess of *Alu* elements. This implies that the ARMD process may contribute to effective removal of *Alu* elements from regions in which they have reached high densities. Given the fact that abnormally high *Alu* density within a particular genomic region would also make it prone to recombination-mediated deletions, this result may reflect a selective force that counteracts the deletion process.

A Potential Mechanism of Double-Strand Break (DSB) Repair

Previous analyses have demonstrated the ability of both LTR and non-LTR retrotransposons to cause DSBs in genomic DNA.[57,58] In particular, the role of the L1 family in the creation and subsequent resolution of DSBs has been extensively analyzed.[43] In vitro, cell-culture studies have shown that homology-directed repair is a major mechanism for patching such breaks and that recombination between repetitive elements is one possible pathway for this process.[59] Recombination rates are highly increased on artificially induced DSBs in cultured cells, which further implicates this mechanism in "tying up the loose ends" at potentially deleterious DSB loci.[60]

In vitro, a 3:1 excess of recombination deletions versus conservative noncrossover situations was detected in a study of homology-mediated repair at a single predefined DSB locus.[60] In this context, some of the loci in our study may represent instances of homology-mediated DSB repair, in which the presence of highly conserved *Alu* sequences on both sides of the break has facilitated its patching. This would be particularly true for loci at which the deletion would otherwise be selectively neutral, since the act of repairing a potentially lethal DSB would give it an instant advantage, if only for propagation to the immediately next generation.

Conclusion

As high-throughput sequencing techniques become more advanced, the focus of evolutionary studies is shifting more toward genomewide analyses. Our study represents such a situation: we have comprehensively analyzed a major deletion mechanism in the human genome that was previously known only as a result of mutations in isolated disease-causing loci. In view of the fact that deletions are being recognized as an important class of genetic variants that contribute to human diversity and evolution,[39–41] ARMD represents one of the major mechanisms for generating such deletions in humans. Moreover, the frequent occurrence of ARMD in gene-rich regions of the genome demonstrates the importance of this process in both biomedical and evolutionary studies. Overall, our results open the field to further studies of deletions caused by recombination between mobile elements and demonstrate one of the possible ways by which the human lineage may have developed a set of unique genetic traits.

Acknowledgments

We thank Drs. J. Xing, M. Konkel, S. W. Herke, and two anonymous reviewers, for their useful comments on earlier versions of the manuscript, and D. Srikanta and L. Song, for technical assistance. We are especially grateful to J. A. Walker, for her help throughout this project, and to Dr. J. Kim, for valuable technical advice. This research was supported by National Science Foundation grants BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.); Louisiana Board of Regents Millennium Trust Health Excellence Fund grants (2000–05)-05 (M.A.B.), (2000–05)-01 (M.A.B.), and (2001–06)-02 (M.A.B.); National Institutes of Health grants RO1 GM59290 (M.A.B.), R03 CA101515 (P.L.), and P30 CA16056 (Roswell Park Cancer Institute); and the State of Louisiana Board of Regents Support Fund (M.A.B.).

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Batzer Laboratory, http://batzerlab.lsu.edu/
Blast 2 Sequences, http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi (for Bl2seq)
DNASTAR, https://www.dnastar.com/web/index.php
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for gorilla and orangutan DNA sequences submitted under accession numbers DQ363502–363524)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *CHRNA9*)
Primer3, http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi
RefSeq, http://www.ncbi.nlm.nih.gov/projects/RefSeq/

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
2. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA (2005) *Alu* retrotransposition-mediated deletion. J Mol Biol 348:791–800
3. Deininger PL, Batzer MA (1999) Alu repeats and human disease. Mol Genet Metab 67:183–193
4. Hattori Y, Okayama N, Ohba Y, Yamashiro Y, Yamamoto K, Tsukimoto I, Kohakura M (1999) The precise breakpoints of a Filipino-type alpha-thalassemia-1 deletion found in two Japanese. Hemoglobin 23:239–248
5. Huang LS, Ripps ME, Korman SH, Deckelbaum RJ, Breslow JL (1989) Hypobetalipoproteinemia due to an apolipoprotein B gene exon 21 deletion derived by Alu-Alu recombination. J Biol Chem 264:11394–11400
6. Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379
7. Levran O, Doggett NA, Auerbach AD (1998) Identification of *Alu*-mediated deletions in the Fanconi anemia gene *FAA*. Hum Mutat 12:145–152
8. Marshall B, Isidro G, Boavida MG (1996) Insertion of a short Alu sequence into the hMSH2 gene following a double cross over next to sequences with chi homology. Gene 174:175–179
9. Myerowitz R, Hogikyan ND (1987) A deletion involving Alu sequences in the β-hexosaminidase α-chain gene of French

Canadians with Tay-Sachs disease. J Biol Chem 262:15396–15399
10. Rohlfs EM, Puget N, Graham ML, Weber BL, Garber JE, Skrzynia C, Halperin JL, Lenoir GM, Silverman LM, Mazoyer S (2000) An *Alu*-mediated 7.1 kb deletion of *BRCA1* exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10. Genes Chromosomes Cancer 28:300–307
11. Rothberg PG, Ponnuru S, Baker D, Bradley JF, Freeman AI, Cibis GW, Harris DJ, Heruth DP (1997) A deletion polymorphism due to Alu-Alu recombination in intron 2 of the retinoblastoma gene: association with human gliomas. Mol Carcinog 19:69–73
12. Tvrdik T, Marcus S, Hou SM, Falt S, Noori P, Podlutskaja N, Hanefeld F, Stromme P, Lambert B (1998) Molecular characterization of two deletion events involving *Alu*-sequences, one novel base substitution and two tentative hotspot mutations in the hypoxanthine phosphoribosyltransferase (HPRT) gene in five patients with Lesch-Nyhan syndrome. Hum Genet 103:311–318
13. Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87
14. Miyamoto MM, Slightom JL, Goodman M (1987) Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. Science 238:369–373
15. Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus *Homo*. Proc Natl Acad Sci USA 100: 7181–7188
16. Deininger PL, Batzer MA (2002) Mammalian retroelements. Genome Res 12:1455–1465
17. Hamaker HC (1978) Approximating the cumulative normal distribution and its inverse. Appl Stat 27:76–77
18. Inoue K, Lupski JR (2002) Molecular mechanisms for genomic disorders. Annu Rev Genomics Hum Genet 3:199–242
19. Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, Walker JA, Watkins WS, Jorde LB, Batzer MA (2004) Genome-wide analysis of the human Alu Yb-lineage. Hum Genomics 1:167–178
20. Otieno AC, Carter AB, Hedges DJ, Walker JA, Ray DA, Garber RK, Anders BA, Stoilova N, Laborde ME, Fowlkes JD, Huang CH, Perodeau B, Batzer MA (2004) Analysis of the human *Alu* Ya-lineage. J Mol Biol 342:109–118
21. Rudiger NS, Gregersen N, Kielland-Brandt MC (1995) One short well conserved region of *Alu*-sequences is involved in human gene rearrangements and has homology with prokaryotic *chi*. Nucleic Acids Res 23:256–260
22. Cordaux R, Lee J, Dinoso L, Batzer MA. Recently integrated *Alu* retrotransposons are essentially neutral residents of the human genome. Gene (http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T39-4JF8HCB-3&_coverDate=03%2F09%2F2006&_alid=397417522&_rdoc=1&_fmt=&_orig=search&_qd=1&_cdi=4941&_sort=d&view=c&_acct=C000001358&_version=1&_urlVersion=0&_userid=5745&md5=1acad274031312a563629c5232294196)(electronically published March 6, 2006; accessed May 2, 2006)
23. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA (2005) Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res 33:4040–4052
24. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945
25. Sakharkar MK, Chow VT, Kangueane P (2004) Distributions of exons and introns in the human genome. In Silico Biol 4: 387–393
26. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al (2001) The sequence of the human genome. Science 291:1304–1351
27. Lustig LR, Peng H (2002) Chromosome location and characterization of the human nicotinic acetylcholine receptor subunit alpha (α) 9 (CHRNA9) gene. Cytogenet Genome Res 98:154–159
28. Brookfield JF (2001) Selection on Alu sequences? Curr Biol 11:R900–901
29. Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV (2004) Duplication, coclustering, and selection of human Alu retrotransposons. Proc Natl Acad Sci USA 101:1268–1272
30. Hackenberg M, Bernaola-Galvan P, Carpena P, Oliver JL (2005) The biased distribution of Alus in human isochores might be driven by recombination. J Mol Evol 60:365–377
31. Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE (2003) Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res 13:358–368
32. Petrov DA (2001) Evolution of genome size: new approaches to an old problem. Trends Genet 17:23–28
33. Bailey JA, Liu G, Eichler EE (2003) An *Alu* transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834
34. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA (2004) Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 14:1068–1075
35. Burwinkel B, Kilimann MW (1998) Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. J Mol Biol 277:513–517
36. Dombroski BA, Scott AF, Kazazian HH Jr (1993) Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. Proc Natl Acad Sci USA 90:6513–6517
37. Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079
38. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949–951
39. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38:75–81
40. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82–85
41. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM, International HapMap Consortium (2006) Common deletion polymorphisms in the human genome. Nat Genet 38: 86–92
42. Gilbert N, Lutz-Prigge S, Moran JV (2002) Genomic deletions created upon LINE-1 retrotransposition. Cell 110:315–325
43. Gilbert N, Lutz S, Morrish TA, Moran JV (2005) Multiple fates

of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol 25:7780–7795
44. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD (2002) Human L1 retrotransposition is associated with genetic instability in vivo. Cell 110:327–338
45. Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. Gene 324:15–34
46. Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70:1490–1497
47. Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, et al (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 429: 382–388
48. Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD, Rocchi M, Eichler EE (2005) A genome-wide survey of structural variation between human and chimpanzee. Genome Res 15:1344–1356
49. Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960–1963
50. Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. Cell 119:1027–1040
51. Babcock M, Pavlicek A, Spiteri E, Kashork CD, Ioshikhes I, Shaffer LG, Jurka J, Morrow BE (2003) Shuffling of genes within low-copy repeats on 22q11 (LCR22) by *Alu*-mediated recombination events during evolution. Genome Res 13:2519–2532
52. Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N (2001) *Alu*-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. Proc Natl Acad Sci USA 98: 11399–11404
53. Szabo Z, Levi-Minzi SA, Christiano AM, Struminger C, Stoneking M, Batzer MA, Boyd CD (1999) Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. J Mol Evol 49:664–671
54. Glowatzki E, Fuchs PA (2000) Cholinergic synaptic inhibition of inner hair cells in the neonatal mammalian cochlea. Science 288:2366–2368
55. Martinez I, Rosa M, Arsuaga JL, Jarabo P, Quam R, Lorenzo C, Gracia A, Carretero JM, Bermudez de Castro JM, Carbonell E (2004) Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. Proc Natl Acad Sci USA 101:9976–9981
56. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of *FOXP2,* a gene involved in speech and language. Nature 418:869–872
57. Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357:1383–1393
58. Zimmerly S, Guo H, Perlman PS, Lambowitz AM (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82:545–554
59. Richardson C, Jasin M (2000) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. Mol Cell Biol 20:9068–9075
60. Liang F, Han M, Romanienko PJ, Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci USA 95:5172–5177